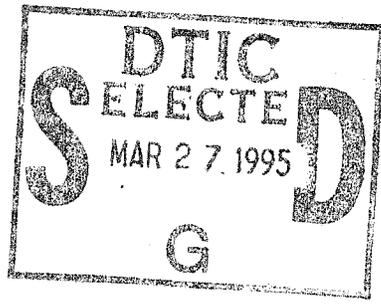




A Page Test With Nuisance Parameter Estimation

Douglas A. Abraham
Submarine Sonar Department



Naval Undersea Warfare Center Division
Newport, Rhode Island

19950323 099

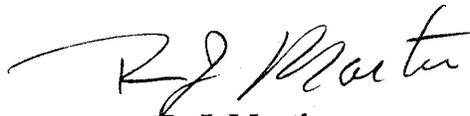
Approved for public release; distribution is unlimited.

PREFACE

The work presented in this report was sponsored by the Naval Undersea Warfare Center's (NUWC's) Independent Research (IR) Program, as Project B10001, "Quickest Detection of Signals in the Presence of Noise with Unknown Parameters." The IR Program is funded by the Office of Naval Research; the NUWC program manager is Dr. Kenneth M. Lima (Code 102).

The author of this report is located at the Naval Undersea Warfare Center Detachment, New London, CT 06320. The technical reviewer for this report was A. H. Nuttall (Code 302). The author thanks Tod Loginbuhl (Code 2121) for posing the problem that led to the development of this algorithm.

Reviewed and Approved: 1 March 1995

A handwritten signature in black ink, appearing to read "R. J. Martin". The signature is fluid and cursive, with the first letters of the first and last names being capitalized and prominent.

R. J. Martin
Acting Head, Submarine Sonar Department

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE 1 March 1995	3. REPORT TYPE AND DATES COVERED Final	
4. TITLE AND SUBTITLE A Page Test With Nuisance Parameter Estimation		5. FUNDING NUMBERS B10001	
6. AUTHOR(S) Douglas A. Abraham			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Undersea Warfare Center Detachment 39 Smith Street New London, Connecticut 06320-5594		8. PERFORMING ORGANIZATION REPORT NUMBER TR 10,829	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Naval Undersea Warfare Center Division 1176 Howell Street Newport, Rhode Island 02841-1708		10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.		12b. DISTRIBUTION CODE	
<p>13. ABSTRACT (Maximum 200 words)</p> <p>The detection of the onset of a signal or the detection of a finite duration signal is a common and relevant problem in sonar signal processing. The Page test using the log-likelihood ratio is the optimal detector structure for signal onset detection when it is desired to minimize the average time before detection (\bar{D}) while constraining the average time between false alarms (\bar{T}). Parameterizations of realistic detection problems typically include unknown parameters having the same value under the signal-present and signal-absent hypotheses, known as nuisance parameters. When testing a finite set of data, nuisance parameters may be dealt with through the use of uniformly most powerful tests, invariant tests, Bayesian approaches, or the generalized likelihood ratio test. Unfortunately, these techniques do not always extend to sequential detection problems. In this report, the Page test is generalized to account for nuisance parameters. The inherent signal-absent decision-making of the Page test is exploited to identify signal-free data (i.e., auxiliary data) to estimate the nuisance parameters. Due to the independence of the auxiliary data and the current Page test statistic, analysis is feasible. Wald- and Siegmund-based approximations to \bar{D} and \bar{T} are derived and shown to simplify to those of the standard Page test when the nuisance parameters are known exactly. Closed forms for the average sample numbers (\bar{D} and \bar{T}) for the Page test with nuisance parameter estimation for a Gaussian shift in mean signal with unknown variance are derived. The approximations are verified through comparison with simulation results, where it is seen, as in the standard Page test, that the Siegmund-based approximation provides more accuracy. It is shown that the linear asymptotic (in the sense of a large threshold) relationship between the threshold and \bar{D} is preserved when the nuisance parameters are estimated. However, the exponential asymptotic relationship between the threshold and \bar{T} of the standard Page test becomes a power law that approximates an exponential for large amounts of auxiliary data. Thus, the loss associated with the estimation of the nuisance parameters is quantified in relation to the amount of auxiliary data available.</p>			
14. SUBJECT TERMS sonar signal processing log-likelihood ratio nuisance parameters Page test Siegmund-based approximation signal onset detection Wald-based approximation			15. NUMBER OF PAGES 48
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR

TABLE OF CONTENTS

	Page
LIST OF ILLUSTRATIONS	ii
LIST OF TABLES	iii
LIST OF SYMBOLS	iv
LIST OF ABBREVIATIONS AND ACRONYMS	v
INTRODUCTION	1
BACKGROUND - THE PAGE TEST	2
AVERAGE SAMPLE NUMBER ANALYSIS	4
EXAMPLES	9
PROPOSED ALGORITHM	16
NUISANCE PARAMETER ESTIMATION	17
AVERAGE SAMPLE NUMBER ANALYSIS	19
GAUSSIAN SHIFT IN MEAN SIGNAL WITH UNKNOWN VARIANCE	24
CONCLUSIONS	36
REFERENCES	37

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

LIST OF ILLUSTRATIONS

Figure		Page
1	Sample Page Test Statistic Sequence	3
2	Simulation and Wald and Siegmund Approximations to the Average Delay Before Detection for Gaussian Shift in Mean Signal With $h = 5$	13
3	Simulation and Wald and Siegmund Approximations to the Average Delay Before Detection for Exponential Signal With $h = 5$	13
4	Simulation and Wald and Siegmund Approximations to the Average Delay Before Detection for Noncentral Chi-Squared Signal With $h = 10$	14
5	Simulation and Wald and Siegmund Approximations to the Average Time Between False Alarms for Gaussian Shift in Mean Signal	14
6	Simulation and Wald and Siegmund Approximations to the Average Time Between False Alarms for Exponential Signal	15
7	Simulation and Wald and Siegmund Approximations to the Average Time Between False Alarms for Noncentral Chi-Squared Signal	15
8	Placement of Auxiliary Data in Relation to the Most Recent Reset of the Page Test Statistic	18
9	Simulation and Wald and Siegmund Approximations to the Average Delay Before Detection for Gaussian Shift in Mean Signal With Threshold $h = 5$ for Page Test With Nuisance Parameter Estimation	31
10	Simulation and Wald and Siegmund Approximations to the Average Time Between False Alarms for Gaussian Shift in Mean Signal for Page Test With Nuisance Parameter Estimation	31
11	Average Delay Before Detection Vs. Average Time Between False Alarms for 0 dB SNR and Varying Amounts of Auxiliary Data	32
12	Average Time Between False Alarms Vs. Threshold for 0 dB SNR and Varying Amounts of Auxiliary Data	32
13	Average Delay Before Detection Vs. SNR for $\bar{T} = 10^4$ Samples and Varying Amounts of Auxiliary Data	33

LIST OF TABLES

Table		Page
1	Siegmund Corrections to Sample Signals	10
2	Siegmund Corrections for Noncentral Chi-Squared Signal for $n = 2$ and Several Values of δ	10
3	Information About Sample Signals	11
4	Mean, Variance, and MGF Unity-Root for Gaussian Shift in Mean Signal Conditioned on the NPE $\hat{\sigma}^2$	25

LIST OF SYMBOLS

$A_S(\theta)$	Siegmund-based approximation to the average number of samples for the Page test with nuisance parameter estimation when θ is the signal strength
$A_W(\theta)$	Wald-based approximation to the average number of samples for the Page test with nuisance parameter estimation when θ is the signal strength
\bar{D}	Average delay before detection for standard Page test
\bar{D}_S	Siegmund-based approximation to average delay before detection for Page test with nuisance parameter estimation
\bar{D}_W	Wald-based approximation to average delay before detection for Page test with nuisance parameter estimation
e	Natural logarithm base $\approx 2.71828\dots$
γ	Magnitude of Siegmund correction for Gaussian shift in mean signal ≈ 0.5825
$g(x)$	Detector nonlinearity
$g_{\hat{\lambda}}(x)$	Detector nonlinearity with nuisance parameter estimation
h	Detector threshold
J	Amount of auxiliary data
λ	Nuisance parameter
$\hat{\lambda}$	Nuisance parameter estimate
$P(\theta)$	Probability that Wald test terminates at zero when θ is the signal strength
ρ_-	Siegmund correction term for excess below zero
ρ_+	Siegmund correction term for excess above h
S_n	Cumulative summation statistic at time index n
S_N	Cumulative summation statistic when test has terminated
\bar{T}	Average time between false alarms for standard Page test
\bar{T}_S	Siegmund-based approximation to average time between false alarms for Page test with nuisance parameter estimation
\bar{T}_W	Wald-based approximation to average time between false alarms for Page test with nuisance parameter estimation
θ	Signal strength parameter
θ_0	Signal strength for signal-absent hypothesis
θ_1	Signal strength for signal-present hypothesis
W_k	Page test statistic

LIST OF ABBREVIATIONS AND ACRONYMS

ASN	Average sample number
LLR	Log-likelihood ratio
MGF	Moment generating function
MLE	Maximum likelihood estimate
NPE	Nuisance parameter estimate
PDF	Probability density function
SLRT	Sequential likelihood ratio test
SNR	Signal-to-noise ratio
SPRT	Sequential probability ratio test

A PAGE TEST WITH NUISANCE PARAMETER ESTIMATION

INTRODUCTION

Signal detectors may be described as parametric, where the observed data are assumed to follow a parameterized probability density function (PDF), or nonparametric. Parametric detectors provide performance improvement over nonparametric detectors at the expense of increased sensitivity to an incorrect parameterization. Many parameterizations of realistic detection problems result in composite hypothesis tests where there are unknown nuisance and signal parameters. Nuisance parameters are unknown parameters having the same value under both the signal-present and signal-absent hypotheses.

When testing a finite set of data, there are several techniques for dealing with nuisance parameters: uniformly most powerful tests, Bayesian approaches when prior distributions on the nuisance parameters are known or can be assumed, invariant tests, the generalized likelihood ratio test, or ad hoc techniques such as the substitution of maximum likelihood estimates (MLEs) of the nuisance parameters from auxiliary data in a log-likelihood ratio (LLR) or locally optimal nonlinearity. Such techniques do not always extend to sequential detection problems. As illustrated by Govindarajulu [1], most sequential tests for composite hypotheses essentially estimate unknown parameters using all data previous to the current sample. The disadvantages of these techniques are that they do not allow nonstationary nuisance parameters and are potentially difficult to analyze. Stahl [2] used a similar procedure in the Page test [3] for a change detection problem where the nuisance parameters were estimated by an exponential averager over all data previous to the current sample. Disadvantages of this technique are the lack of theoretical analysis of the performance and the corruption of the nuisance parameter estimate (NPE) by data containing both signal and noise. This report proposes and analyzes a method for the estimation of nuisance parameters in the Page test for the change detection problem.

BACKGROUND - THE PAGE TEST

The change detection problem, in general, may be characterized by a model assuming that the observed data

$$x_1, x_2, x_3, \dots \quad (1)$$

are statistically independent and distributed according to the rule

$$x_i \sim \begin{cases} f(x|\lambda, \theta_0) & i < p \\ f(x|\lambda, \theta_1) & i \geq p \end{cases}, \quad (2)$$

where $f(x|\lambda, \theta)$ is a parameterized PDF having λ as an unknown nuisance parameter and θ as a signal strength parameter. Here, θ_0 is the value of θ when no signal is present (typically $\theta_0 = 0$), θ_1 is the value of θ when signal is present, and p is the unknown starting time of the signal.

The change detection problem amounts to determining, as rapidly as possible, when the sequence of data changes from following the signal-absent probability distribution law to following the signal-present law. If $g(x)$ is the detector nonlinearity applied to the observed data, the Page test declares a detection when the Page test statistic,

$$W_k = \max \{0, W_{k-1} + g(x_k)\}, \quad k \geq 1, \quad (3)$$

is greater than a threshold h , where $W_0 = 0$. A sample Page test statistic sequence is found in figure 1.

As with most sequential tests, the primary performance measures for the Page test are the average number of samples* before a threshold crossing under the signal-present and signal-absent hypotheses. For the Page test, these are respectively called the average number of samples before a detection, \bar{D} , and the average number of samples between or before a false alarm, \bar{T} .

*Also known as the average sample number, average stopping time, or average record length.

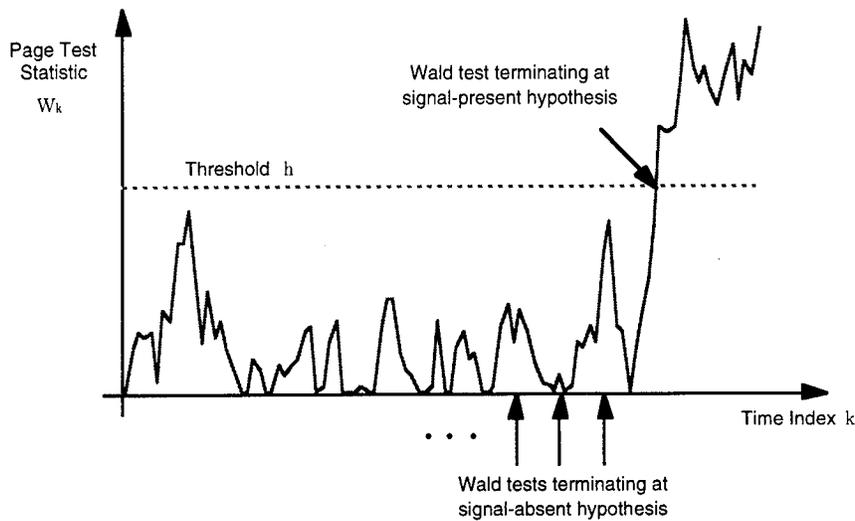


Figure 1. Sample Page Test Statistic Sequence

When $g(x)$ is the LLR of a data sample, the Page test yields the smallest worst case \bar{D} when \bar{T} is lower bounded by \bar{T}_0 . The worst case \bar{D} is observed when the signal starts at a regulation or reset of the Page test statistic (i.e., $W_k = 0$). Had the signal started at some other time index, the Page test statistic may have crossed the threshold sooner, resulting in a smaller value for \bar{D} . The optimality was shown asymptotically as $\bar{T}_0 \rightarrow \infty$ by Lorden [4] and subsequently for finite \bar{T}_0 by Moustakides [5].

If the signal-present hypothesis is characterized by an unknown signal strength, the locally optimal nonlinearity may be substituted for the LLR in $g(x)$. Dyson [6] showed that this results in the same asymptotic optimality (i.e., Lorden's result) as the LLR in the Page test as the signal strength goes to zero.

The Page test has also been applied to the detection of a temporary change in the PDF of the observed data. In this case, the false alarm performance measure (\bar{T}) remains the same. However, the detection performance is characterized by the probability of detecting the signal. Analytical methods for approximating the probability of detection that are extremely accurate when $g(x_k)$ is Gaussian are discussed by Han, Willett, and Abraham [7]. A simulation based method providing accurate approximation for non-Gaussian Page test updates may be found in [8].

AVERAGE SAMPLE NUMBER ANALYSIS

Basseville and Nikiforov [9] discuss the primary analytical results for determining the average sample numbers (ASNs) for the Page test. These are based on representing the Page test as consecutive Wald tests terminating at the signal-absent or null hypothesis, followed by a final Wald test terminating at the signal-present hypothesis. The termination of the Wald tests at the null hypothesis is reflected in the Page test statistic as a regulation or reset to zero.

A Wald test may be described by the stopping time

$$N = \inf \{n > 0 : S_n \notin (a, b)\} \quad (4)$$

where a and b are thresholds chosen to satisfy Type I and II error probabilities and

$$S_n = \sum_{i=1}^n g(x_i) \quad (5)$$

is the cumulative summation of the data $\{x_i\}_{i=1}^n$ transformed by the nonlinearity g . When $g(x)$ is the LLR, the test is known as the sequential likelihood ratio test (SLRT) or sequential probability ratio test (SPRT) and is optimal in the sense of providing the smallest ASNs of any test with equivalent Type I and II error probabilities [10].

The ASN analysis of the Page test employs Wald's first equation ([11], [9], [12])

$$E[N] = \frac{E[S_N]}{E[g(x_i)]}, \quad (6)$$

and Wald's fundamental identity

$$E[e^{tS_N} M(t)^{-N}] = 1, \quad (7)$$

where

$$M(t) = E[e^{t g(x_i)}] \quad (8)$$

is the moment generating function (MGF) of the transformed data. Equation (6) is used to approximate the average number of samples for the Wald test terminating at either the signal-present or signal-absent hypotheses, and equation (7) is used to approximate the

probability that the Wald test terminates at the signal-absent hypothesis as a function of the signal strength θ (also known as the operating characteristic of the Wald test). How the approximations are carried out forms the differences between the techniques discussed by Basseville and Nikiforov [9].

Wald Approximation

The first, less accurate result utilizes Wald's approximations to the ASNs and operating characteristic, $P(\theta)$, of a Wald test. They are based on the assumption that when the test stops, the value of the test statistic is close to the value of the threshold; that is, $S_N \approx a$ or $S_N \approx b$. As seen in Basseville and Nikiforov [9], this method leads to the approximations for the Page test ASNs

$$\bar{T} \approx \frac{1 + ht_{\theta_0} - e^{ht_{\theta_0}}}{t_{\theta_0} E_{\theta_0} [g(x_i)]} \quad (9)$$

and

$$\bar{D} \approx \frac{1 + ht_{\theta_1} - e^{ht_{\theta_1}}}{t_{\theta_1} E_{\theta_1} [g(x_i)]}, \quad (10)$$

where the expectations are taken, respectively, under the signal-absent (θ_0) and signal-present (θ_1) hypotheses. Here, t_θ , for $\theta = \theta_0$ or θ_1 , is the nonzero unity-root of the MGF

$$E_\theta [e^{t_\theta g(x_i)}] = M(t_\theta) = 1. \quad (11)$$

Siegmund Approximation

The second approximation discussed by Basseville and Nikiforov [9] utilizes corrections, derived by Siegmund [11], to the Wald approximations to the ASNs and operating characteristic of the Wald test. The corrections attempt to account for the excess above or below the threshold in the Wald test at termination. The resulting approximations to \bar{T} and \bar{D} ([11], [9]) are identical in form to those of equations (9) and (10) with the threshold h replaced by

$$h_S = h + \rho_+ - \rho_- \quad (12)$$

where ρ_+ and ρ_- represent the corrections due to excess beyond a threshold.

The corrections are derived, as seen in Siegmund [11], by first normalizing the Wald test statistic to have zero-mean and unit-variance. If the mean and variance of $g(x)$ are μ and σ^2 respectively, then the normalized update is

$$Y_i = \frac{g(x_i) - \mu}{\sigma} . \quad (13)$$

Let the cumulative summation be described by

$$V_n = \sum_{i=1}^n Y_i \quad (14)$$

and let the test (that stops when the upper threshold h is crossed) have the stopping time

$$N_+ = \inf \{n > 0 : V_n \geq h\} . \quad (15)$$

A lemma described by Siegmund [13] states that the excess beyond the threshold in the Wald test statistic at termination may be approximated by

$$\mathbb{E} [V_{N_+} - h] \approx \frac{\mathbb{E} [V_{N_+}^2]}{2\mathbb{E} [V_{N_+}]} . \quad (16)$$

In [11], Siegmund shows that

$$\frac{\mathbb{E} [V_{N_+}^2]}{2\mathbb{E} [V_{N_+}]} = \frac{\kappa}{6} - \frac{H}{\pi} , \quad (17)$$

where

$$H = \int_{\lambda=0}^{\infty} \lambda^{-2} \text{Re} \log \left\{ \frac{2}{\lambda^2} [1 - \Phi(\lambda)] \right\} d\lambda , \quad (18)$$

$$\Phi(\lambda) = \mathbb{E} [e^{j\lambda Y_i}] \quad (19)$$

is the characteristic function of Y_i ,

$$\begin{aligned} \kappa &= \mathbb{E} [Y_i^3] \\ &= \frac{\mathbb{E} [(g(x_i) - \mu)^3]}{\sigma^3} \end{aligned} \quad (20)$$

is the third moment of Y_i or, equivalently, the skewness of $g(x_i)$, and the expectations are taken under either the signal-present or signal-absent hypotheses, depending on whether the correction is for \bar{D} or \bar{T} .

The excess beyond the lower threshold will simply be the negative of equation (17) if the Wald test update has a symmetric distribution as is the case with a Gaussian $g(x_i)$ (this is all that Siegmund considers in [11]). Let the test that stops when the lower threshold is crossed have the stopping time

$$N_- = \inf \{n > 0 : V_n \leq 0\} . \quad (21)$$

The excess in the Wald test statistic at termination may then be approximated by

$$\begin{aligned} E[V_{N_-}] &\approx \frac{E[V_{N_-}^2]}{2E[V_{N_-}]} \\ &= -\frac{E[(-V_{N_-})^2]}{2E[-V_{N_-}]} \\ &= -\frac{E[W_{M_+}^2]}{2E[W_{M_+}]} , \end{aligned} \quad (22)$$

where

$$W_n = -V_n = \sum_{i=1}^n (-Y_i) \quad (23)$$

and

$$M_+ = \inf \{n > 0 : W_n \geq 0\} = N_- , \quad (24)$$

which is the same form as the above (equations (14)–(16)) except that Y_i is replaced by $-Y_i$. Noting that the third central moment of $-Y_i$ is simply $-\kappa$ and that the characteristic function of $-Y_i$ is $\Phi(-\lambda)$, equation (22) may be simplified, using equation (17), to

$$-\frac{E[W_{M_+}^2]}{2E[W_{M_+}]} = \frac{\kappa}{6} + \pi^{-1} \int_{\lambda=0}^{\infty} \lambda^{-2} \operatorname{Re} \log \left\{ \frac{2}{\lambda^2} [1 - \Phi(-\lambda)] \right\} d\lambda . \quad (25)$$

The natural logarithm of a complex number ($z = re^{j\theta}$) may be described in Cartesian form through the following identity

$$\begin{aligned} \log z &= \log re^{j\theta} \\ &= \log r + j\theta . \end{aligned} \quad (26)$$

Thus, the real part of the natural logarithm of a complex number is simply the logarithm of the magnitude,

$$\operatorname{Re} \log z = \log r . \quad (27)$$

Use of this identity in equation (18) results in the simplification

$$H = \int_{\lambda=0}^{\infty} \lambda^{-2} \log \left\{ \frac{2}{\lambda^2} |1 - \Phi(\lambda)| \right\} d\lambda . \quad (28)$$

Now, using the identity

$$\begin{aligned} \Phi(\lambda) &= \mathbf{E} [e^{j\lambda Y_i}] \\ &= \mathbf{E} [\cos(\lambda Y_i)] + j\mathbf{E} [\sin(\lambda Y_i)] , \end{aligned} \quad (29)$$

it can be shown that

$$\begin{aligned} |1 - \Phi(\lambda)| &= |1 - \mathbf{E} [\cos(\lambda Y_i)] - j\mathbf{E} [\sin(\lambda Y_i)]| \\ &= \left\{ (1 - \mathbf{E} [\cos(\lambda Y_i)])^2 + \mathbf{E} [\sin(\lambda Y_i)]^2 \right\}^{\frac{1}{2}} \\ &= \left\{ (1 - \mathbf{E} [\cos(-\lambda Y_i)])^2 + \mathbf{E} [\sin(-\lambda Y_i)]^2 \right\}^{\frac{1}{2}} \\ &= |1 - \Phi(-\lambda)| \end{aligned} \quad (30)$$

is an even function of λ . Thus, equation (25) simplifies to

$$-\frac{\mathbf{E} [W_{M+}^2]}{2\mathbf{E} [W_{M+}]} = \frac{\kappa}{6} + \frac{H}{\pi} . \quad (31)$$

Incorporating a non-unit-variance into the Wald test update simply results in scaling equations (17) and (31) by the standard deviation σ to yield the Siegmund corrections

$$\rho_+ = \sigma \left\{ \frac{\kappa}{6} - \frac{H}{\pi} \right\} \quad (32)$$

and

$$\rho_- = \sigma \left\{ \frac{\kappa}{6} + \frac{H}{\pi} \right\} . \quad (33)$$

EXAMPLES

Gaussian shift in mean, exponential, and noncentral chi-squared signal types often arise in sonar signal processing applications. For example, the Gaussian shift in mean signal may represent a range-induced bias in the innovations sequence of a tracking algorithm. The exponential signal type is encountered as the magnitude-squared-basebanded-matched-filter output or as the result of frequency domain processing on a sinusoidal signal with Gaussian random amplitude and uniform random phase corrupted by Gaussian noise (e.g., multiple point reflectors with none dominant for active sonar). The noncentral chi-squared signal is the result of frequency domain processing on a sinusoidal signal with deterministic amplitude and uniform random phase corrupted by Gaussian noise (e.g., a single dominant reflector for active sonar). In the following sections, the Siegmund correction terms for each of these signal types are presented and used to compare the Siegmund and Wald approximations to \bar{D} and \bar{T} to the ASN obtained through simulation.

Siegmund Correction Terms

This section contains the Siegmund correction terms and the information required to obtain them for Gaussian shift in mean, exponential, and noncentral chi-squared signals. Table 1 contains the Siegmund correction terms and the mean, variance, skewness, and characteristic function of the normalized statistic of equation (13) for each of the assumed distributions for the Page test update $g(x)$. Here, $\mathcal{N}(\mu, \sigma^2)$ represents a Gaussian random variable with mean μ and variance σ^2 , $\text{Exp}(\sigma)$ represents an exponential random variable with mean σ , and $\mathcal{X}_n^2(\delta)$ represents a noncentral chi-squared random variable with n degrees of freedom and noncentrality parameter δ .

The results for the Gaussian shift in mean signal may be found in Siegmund [11] or Basseville and Nikiforov [9]. The skewness values and the characteristic functions for the unnormalized random variables may be found in Manoukian [12] or Johnson and Kotz [14]. The Siegmund corrections for the noncentral chi-squared signal, ρ_+^* and ρ_-^* , are functions of n and δ and must be computed individually using equations (32) and (33). Table 2 contains numerical values of the correction terms for $n = 2$ and several values of δ . Note that when

$n = 2$ and $\delta = 0$, a central chi-squared random variable results and the correction terms are identical to those for the exponential signal with $\sigma = 2$, which correctly reflects the relationship between these two types of random variables.

Table 1. Siegmund Corrections to Sample Signals

$g(x)$	$E[g]$	$Var[g]$	κ	$\Phi(\lambda)$	ρ_+	ρ_-
$\mathcal{N}(\mu, \sigma^2)$	μ	σ^2	0	$e^{-\frac{\lambda^2}{2}}$	0.5825σ	-0.5825σ
$\text{Exp}(\sigma)$	σ	σ^2	2	$e^{-j\lambda}(1-j\lambda)^{-1}$	σ	$-\frac{1}{3}\sigma$
$\chi_n^2(\delta)$	$n + \delta$	$\sigma^2 = 2(n + 2\delta)$	$\frac{\sqrt{8(n+3\delta)}}{(n+2\delta)^{\frac{3}{2}}}$	$\frac{\exp\left\{-j\frac{n\lambda}{\sigma} - \frac{2\delta\lambda^2}{\sigma(\sigma-2j\lambda)}\right\}}{(1-j\frac{2\lambda}{\sigma})^{\frac{n}{2}}}$	ρ_+^*	ρ_-^*

Table 2. Siegmund Corrections for Noncentral Chi-Squared Signal for $n = 2$ and Several Values of δ

δ	ρ_+^*	ρ_-^*
0	2	$-\frac{2}{3}$
0.1	2.0948	-0.7008
1	2.6528	-0.9861
10	4.9129	-2.9735

Simulation Comparison

Ideally, the nonlinearity applied to the Page test is the LLR or the locally optimal nonlinearity of the observed data. For the Gaussian shift in mean and exponential signal types, the LLR and locally optimal nonlinearity have the same form: scaled shifted versions of the data. The LLR for the noncentral chi-squared signal has a complicated form [14]; however,

the locally optimal nonlinearity is simply the shifted data. The amount of the shift may be chosen to maximize the asymptotic performance of the Page test for a specified signal-to-noise ratio (SNR) as described in [15]. Table 3 contains the statistical description, LLR or locally optimal nonlinearity, bias term, mean, variance, and SNR definition for each of the signal types.

Table 3. Information About Sample Signals

Type	$g(x)$	Bias (τ)	$E_s[g]$	$\text{Var}_s[g]$	SNR
$\mathcal{N}(\mu, \sigma^2)$	$\frac{\mu}{\sigma^2}x - \tau$	$\frac{\mu^2}{2\sigma^2}$	$\frac{\mu^2}{2\sigma^2}$	$\frac{\mu^2}{\sigma^2}$	$\frac{\mu^2}{\sigma^2}$
$\text{Exp}(1+s)$	$\frac{s}{s+1}x - \tau$	$\log(1+s)$	$s - \tau$	s^2	s
$\chi_2^2(\delta)$	$x - \tau$	$2\left(1 + \frac{2}{\delta}\right)\log\left(1 + \frac{\delta}{2}\right)$	$\delta + 2 - \tau$	$4(\delta + 1)$	δ

In order to determine the Wald and Siegmund approximations to \bar{D} and \bar{T} , the unity-root of the MGF (11) is required under the signal-present and signal-absent hypotheses. As the nonlinearity for the Gaussian shift in mean and exponential signals is their respective LLR, the MGF roots are $t_{\theta_0} = 1$ and $t_{\theta_1} = -1$. The MGF roots for the locally optimal nonlinearity for the noncentral chi-squared signal are the solutions to

$$\frac{\delta t_{\theta_1}}{1 - 2t_{\theta_1}} - \tau t_{\theta_1} - \log(1 - 2t_{\theta_1}) = 0 \quad (34)$$

and

$$\tau t_{\theta_0} + \log(1 - 2t_{\theta_0}) = 0. \quad (35)$$

The Wald and Siegmund approximations for the average delay before detection and that estimated from simulations are found in figures 2–4 for each signal type as a function of SNR. The threshold was $h = 5$ for the Gaussian shift in mean and exponential signals and $h = 10$ for the noncentral chi-squared signal. The estimated value of \bar{D} is the sample mean of the number of samples required for detection, taken over 1000 trials. Observe that for each signal type, the Siegmund approximation is closer to the simulated value than

the Wald approximation, particularly for large SNR. This result is expected, as the Wald approximation ignores the excess over the threshold, which may be substantial for large strength signals.

The Wald and Siegmund approximations for the average time between false alarms and that estimated from simulations are found in figures 5–7 for each signal type as a function of threshold. The nonlinearity for each signal was formed assuming that the signal strength was zero decibels ($\frac{\mu^2}{\sigma^2} = s = \delta = 1$). As with the results for \bar{D} , utilizing the Siegmund correction terms provides a better approximation than the straight Wald method. The quality of the Wald and Siegmund approximations is not as good for \bar{T} as for \bar{D} . This is a direct result of the increased number of resets or regulations of the Page test when no signal is present (i.e., more Wald tests terminating at the signal-absent hypothesis), which introduces a higher sensitivity to errors in the approximations to the ASNs and operating characteristic of the Wald test.

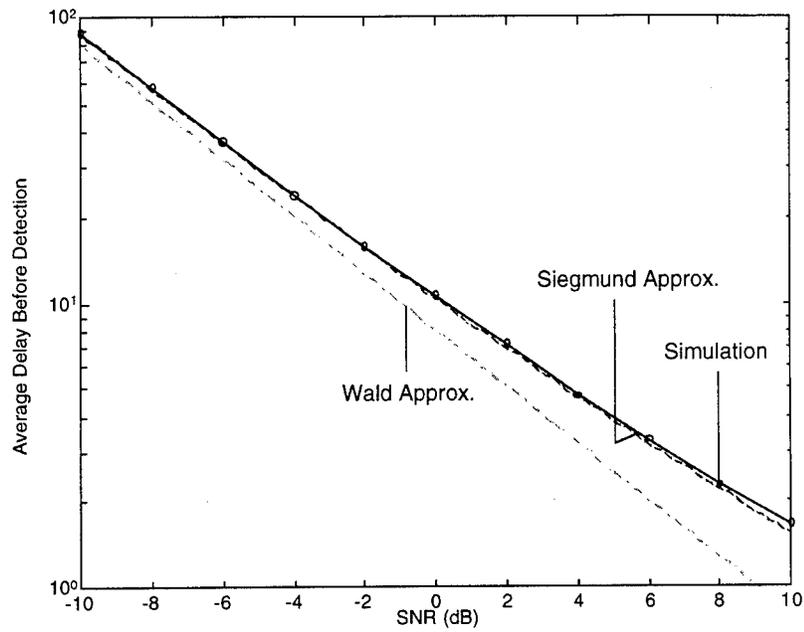


Figure 2. Simulation and Wald and Siegmund Approximations to the Average Delay Before Detection for Gaussian Shift in Mean Signal With $h = 5$

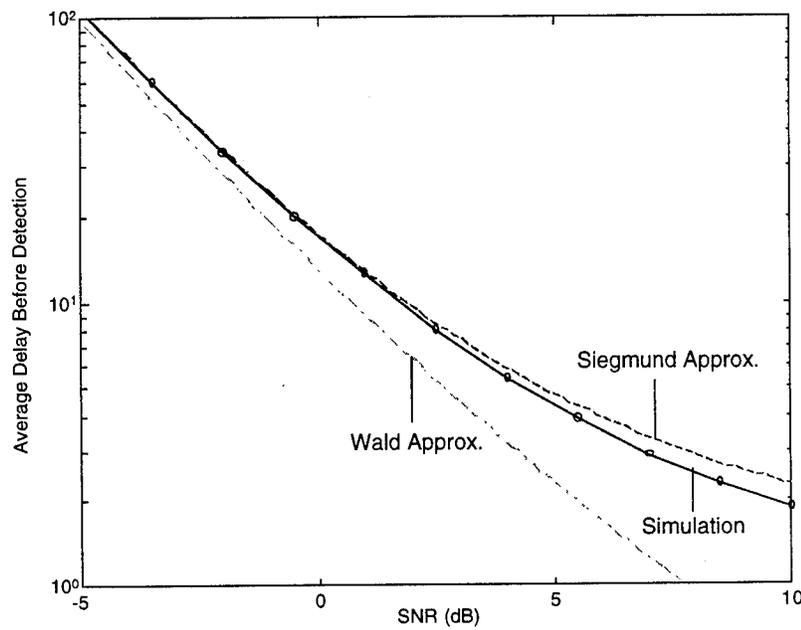


Figure 3. Simulation and Wald and Siegmund Approximations to the Average Delay Before Detection for Exponential Signal With $h = 5$

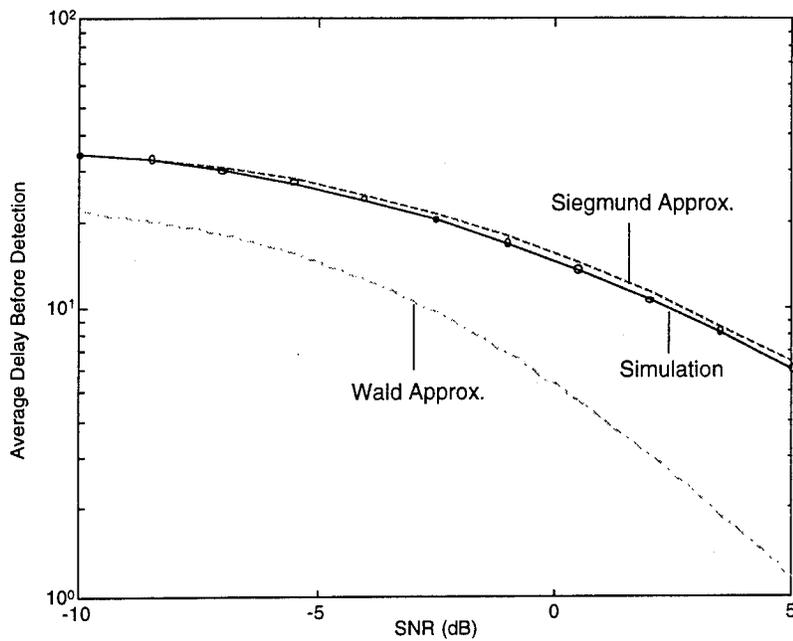


Figure 4. Simulation and Wald and Siegmund Approximations to the Average Delay Before Detection for Noncentral Chi-Squared Signal With $h = 10$

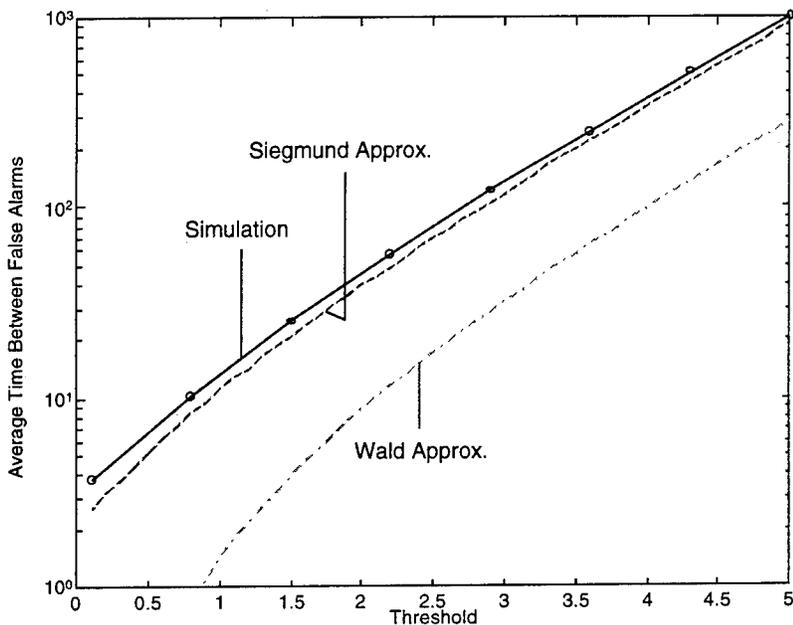


Figure 5. Simulation and Wald and Siegmund Approximations to the Average Time Between False Alarms for Gaussian Shift in Mean Signal

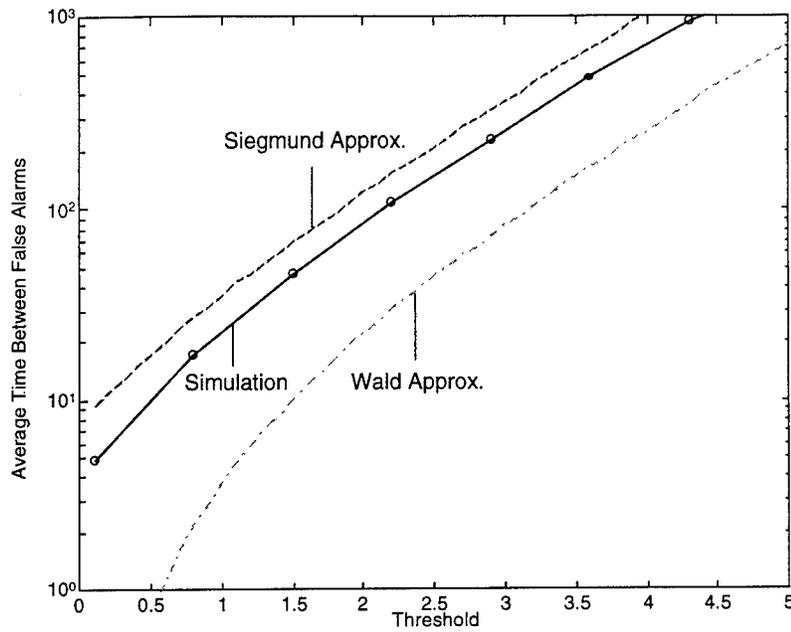


Figure 6. Simulation and Wald and Siegmund Approximations to the Average Time Between False Alarms for Exponential Signal

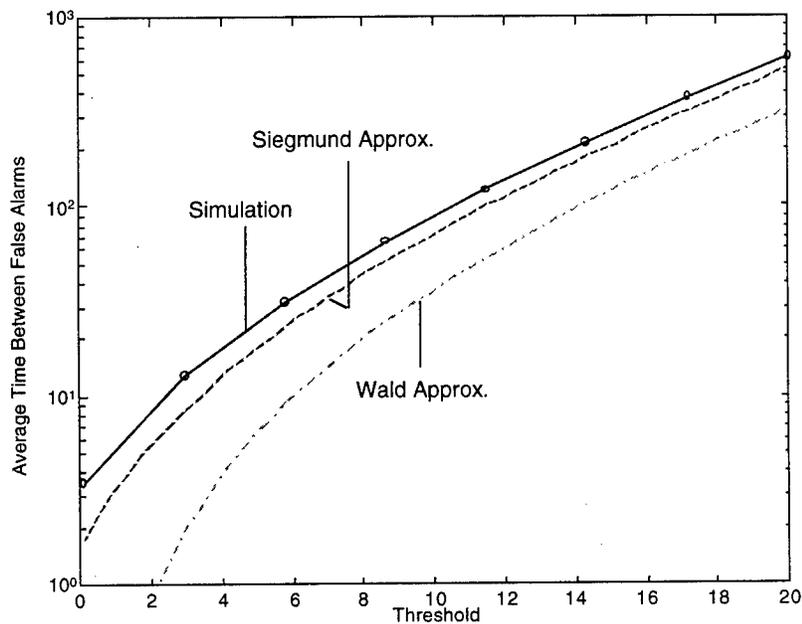


Figure 7. Simulation and Wald and Siegmund Approximations to the Average Time Between False Alarms for Noncentral Chi-Squared Signal

PROPOSED ALGORITHM

As previously mentioned, the Page test [3] may be described as consecutive Wald tests terminating at the signal-absent hypothesis followed by a final Wald test terminating at the signal-present hypothesis. Each termination of a Wald test at the signal-absent hypothesis (represented by a regulation or reset of the Page test to zero) indicates that all previous data are considered to be signal free. Thus, these data may be used to estimate any unknown parameters that also exist under the signal-absent hypothesis; for example, nuisance parameters. An additional benefit of using the data previous to the most recent reset of the Page test is the statistical independence of these data and the Page test statistic after the reset. Thus, analysis may be performed by conditioning on the observed NPE, followed by an expectation removing the conditioning.

The following update algorithm describes the proposed Page test with nuisance parameter estimation. Here, auxiliary data are data that have been determined to be signal-free by the Page test, $\hat{\lambda}$ is the NPE formed from the auxiliary data, and $g_{\hat{\lambda}}(x)$ is the detector nonlinearity.

1. If $W_k = 0$
 - Store new auxiliary data
 - Update $\hat{\lambda}$ using new auxiliary data
2. Set $W_{k+1} = \max \{0, W_k + g_{\hat{\lambda}}(x_{k+1})\}$
3. If $W_{k+1} \geq h$
 - Declare signal present and stop
- Else
 - Set $k = k + 1$ and return to step 1

In the following sections, details associated with the nuisance parameter estimation are discussed and the Siegmund- and Wald-based approximations to the ASN of the Page test with nuisance parameter estimation are derived and shown to simplify to those for the standard Page test when the nuisance parameter is known exactly. The Page test with nuisance

parameter estimation is then applied to detect a Gaussian shift in mean signal where \bar{D} and \bar{T} are derived and compared to simulation results. The Siegmund-based approximations are used to illustrate the Page test performance as a function of the threshold, SNR, and amount of auxiliary data. Finally, the Wald-based approximations are used to illustrate the asymptotic (in the sense of large threshold) relationship between the threshold and \bar{T} and \bar{D} .

NUISANCE PARAMETER ESTIMATION

The NPE is formed from auxiliary data and, depending on the situation, may only be a consistent estimator under the signal-absent hypothesis. Thus, it is important that the auxiliary data truly be signal-free. A necessary but not sufficient condition for ensuring this is to have the auxiliary data consist of data previous to the most recent reset of the Page test. However, due to the inherent delay observed between the time a signal starts and a threshold crossing indicating a detection, there is a nonzero probability that this auxiliary data may contain both signal and noise. To counter this possibility, it is suggested that a fixed number of either samples or resets be used as a buffer zone. Figure 8 depicts these two arrangements.

For extremely weak signals, the delay before detection can be substantial. This may cause the NPE to be corrupted by data containing noise and signal even with a carefully designed buffer zone. This problem may be avoided altogether if there exists an estimator for the nuisance parameter that has the same form under both the signal-absent and signal-present hypotheses.

Most practical scenarios will involve nuisance parameters that are slowly changing in time. In this situation, the auxiliary data should consist of a block of data—as opposed to using all previous data to estimate the nuisance parameter. The size of the block and its positioning should be chosen to ensure that the nuisance parameter is stationary from the beginning of the block of auxiliary data to either a threshold crossing indicating a detection or the next reset of the Page test statistic.

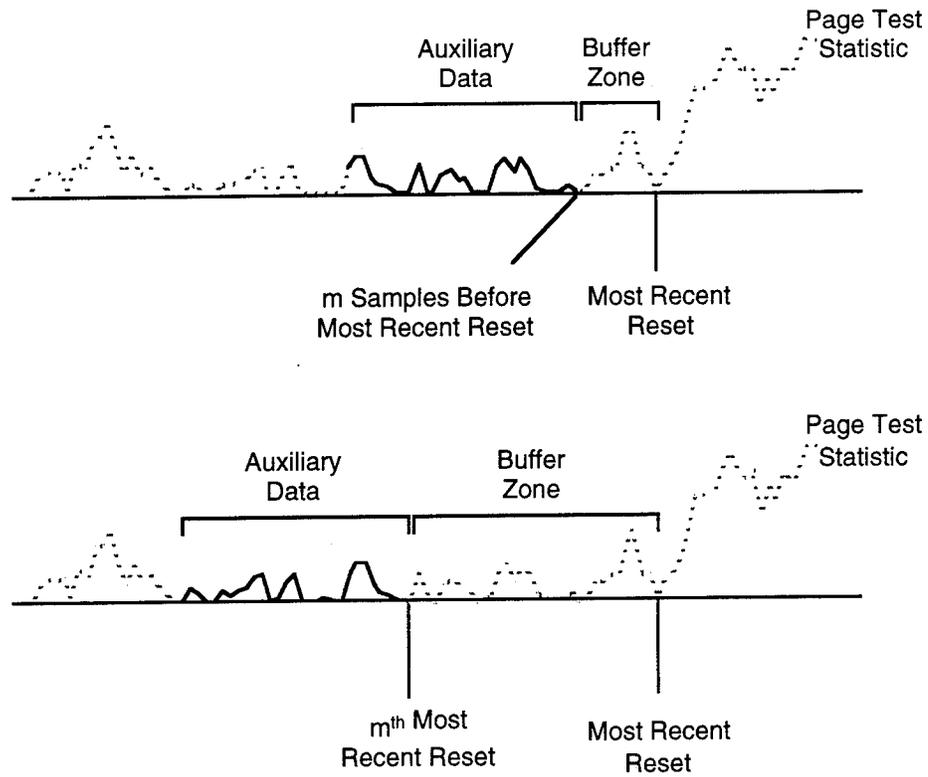


Figure 8. Placement of Auxiliary Data in Relation to the Most Recent Reset of the Page Test Statistic

AVERAGE SAMPLE NUMBER ANALYSIS

As seen in Basseville and Nikiforov [9], the first step in determining the ASN for the Page test is to determine the relationship to the ASNs and operating characteristic for the individual Wald tests that form the Page test. The following definitions are required:

$A(\theta)$ - Average number of samples for the Page test with nuisance parameter estimation when θ is the signal strength

$B_h(\theta)$ - Average number of samples for the Wald test terminating at threshold h when θ is the signal strength

$B_0(\theta)$ - Average number of samples for the Wald test terminating at threshold 0 when θ is the signal strength

$P(\theta)$ - Probability that the Wald test terminates at threshold 0 when θ is the signal strength

Suppose the number of Wald tests that comprise the Page test is m ; that is, $m - 1$ Wald tests terminating at threshold 0 followed by one terminating at threshold h . Then, the ASN for the Page test conditioned on m is

$$A(\theta|m) = (m - 1) B_0(\theta) + B_h(\theta) . \quad (36)$$

The conditioning on m may be removed by taking the expected value

$$\begin{aligned} A(\theta) &= E[A(\theta|m)] \\ &= E[m - 1] B_0(\theta) + B_h(\theta) \\ &= E[m] B_0(\theta) + B_h(\theta) - B_0(\theta) . \end{aligned} \quad (37)$$

Each consecutive Wald test in the Page test is independent of the others. Thus, as found in Manoukian [12], m follows a geometric probability mass function with parameter $q = 1 - P(\theta)$ and mean

$$\begin{aligned} E[m] &= \frac{1}{q} \\ &= \frac{1}{1 - P(\theta)} . \end{aligned} \quad (38)$$

This results in the form

$$A(\theta) = \frac{B_0(\theta)}{1 - P(\theta)} + B_h(\theta) - B_0(\theta) \quad (39)$$

for the ASN of the Page test.

When the Page test includes nuisance parameter estimation, with the update scheme described in the previous section, the ASNs and operating characteristic of the Wald test must be determined conditioned on the NPE. Here, it is assumed that each NPE is independent of the data in the nearby Wald tests and independent of all the other NPEs. This assumption will be valid at least for large buffer sizes and short block sizes. Now, the ASNs and operating characteristic of the Wald test may be described as

$$B_h(\theta) = E_{\hat{\lambda}} [B_h(\theta | \hat{\lambda})] , \quad (40)$$

$$B_0(\theta) = E_{\hat{\lambda}} [B_0(\theta | \hat{\lambda})] , \quad (41)$$

and

$$P(\theta) = E_{\hat{\lambda}} [P(\theta | \hat{\lambda})] , \quad (42)$$

where the arguments of the expectations are the respective ASN or operating characteristic conditioned on the NPE $\hat{\lambda}$.

The test statistic of the Wald test used to implement the Page test at the n^{th} sample is the cumulative summation

$$S_n = \sum_{i=1}^n g_{\hat{\lambda}}(x_i) . \quad (43)$$

Let N be the stopping time of this particular Wald test (i.e., the test has an upper threshold of h and a lower threshold of zero),

$$N = \inf \{n > 0 : S_n \notin (0, h)\} . \quad (44)$$

As found in Siegmund [11] or [13], the average excess of the test statistic beyond the thresholds at termination may be approximated by

$$E [S_N - h | S_N \geq h, \hat{\lambda}] \approx \rho_+ \quad (45)$$

and

$$E[S_N | S_N \leq 0, \hat{\lambda}] \approx \rho_-, \quad (46)$$

where ρ_+ and ρ_- are determined in equations (32) and (33) and are functions of the NPE $\hat{\lambda}$.

Then, using Wald's first equation (6), the ASNs of the Wald test, conditioned on the NPE $\hat{\lambda}$, may be written as

$$\begin{aligned} B_h(\theta | \hat{\lambda}) &= \frac{E[S_N | S_N \geq h, \hat{\lambda}]}{E[g_{\hat{\lambda}}(x) | \hat{\lambda}]} \\ &= \frac{h + E[S_N - h | S_N \geq h, \hat{\lambda}]}{E[g_{\hat{\lambda}}(x) | \hat{\lambda}]} \\ &\approx \frac{h + \rho_+}{E[g_{\hat{\lambda}}(x) | \hat{\lambda}]} \end{aligned} \quad (47)$$

and

$$\begin{aligned} B_0(\theta | \hat{\lambda}) &= \frac{E[S_N | S_N \leq 0, \hat{\lambda}]}{E[g_{\hat{\lambda}}(x) | \hat{\lambda}]} \\ &\approx \frac{\rho_-}{E[g_{\hat{\lambda}}(x) | \hat{\lambda}]} \end{aligned} \quad (48)$$

Following Siegmund [11] or Basseville and Nikiforov [9], the operating characteristic of the Wald test may similarly be related to ρ_+ and ρ_- , except that here it is conditioned on the NPE $\hat{\lambda}$. Let the conditional MGF of the detector nonlinearity $g_{\hat{\lambda}}(x)$ be

$$M_{\hat{\lambda}}(t) = E[e^{t g_{\hat{\lambda}}(x)} | \hat{\lambda}]. \quad (49)$$

Under very mild restrictions on $g_{\hat{\lambda}}(x)$, it can be shown that there exists a nonzero value t_{θ} such that the MGF is one, $M_{\hat{\lambda}}(t_{\theta}) = 1$, where θ is the signal strength parameter. Substitution of this value into Wald's fundamental identity,

$$E[e^{t S_N} M_{\hat{\lambda}}(t)^N | \hat{\lambda}] = 1 \quad (50)$$

results in

$$\begin{aligned} 1 &= E[e^{t_{\theta} S_N} | \hat{\lambda}] \\ &= [1 - P(\theta | \hat{\lambda})] E[e^{t_{\theta} S_N} | S_N \geq h, \hat{\lambda}] + P(\theta | \hat{\lambda}) E[e^{t_{\theta} S_N} | S_N \leq 0, \hat{\lambda}]. \end{aligned} \quad (51)$$

Solving for $1 - P(\theta|\hat{\lambda})$ results in

$$\begin{aligned}
 1 - P(\theta|\hat{\lambda}) &= \frac{1 - E[e^{t\theta} S_N | S_N \leq 0, \hat{\lambda}]}{E[e^{t\theta} S_N | S_N \geq h, \hat{\lambda}] - E[e^{t\theta} S_N | S_N \leq 0, \hat{\lambda}]} \\
 &= \frac{1 - E[e^{t\theta} S_N | S_N \leq 0, \hat{\lambda}]}{e^{h t\theta} E[e^{t\theta} (S_N - h) | S_N \geq h, \hat{\lambda}] - E[e^{t\theta} S_N | S_N \leq 0, \hat{\lambda}]} . \quad (52)
 \end{aligned}$$

Siegmund [11] and Basseville and Nikiforov [9], in analyzing the standard Page test, now proceed by liberally applying a first order approximation to the exponential function. The same results may be obtained by recognizing that the expectations in the numerator and denominator of equation (52) are describable as the MGF of a random variable evaluated at a small argument ($-S_N$ is expected to be small when the test terminates by crossing the threshold at zero and $S_N - h$ is expected to be small when the test terminates by crossing the threshold at h). Thus, the following first order approximation may be applied,

$$\begin{aligned}
 E[e^{aY}] &= E\left[1 + aY + \frac{(aY)^2}{2!} + \dots\right] \\
 &= 1 + aE[Y] + \frac{a^2 E[Y^2]}{2!} + \dots \\
 &\approx 1 + aE[Y] \\
 &\approx e^{aE[Y]} . \quad (53)
 \end{aligned}$$

Disregarding terms on the order of $(t\theta\rho_-)^2$, equation (52) becomes

$$\begin{aligned}
 1 - P(\theta|\hat{\lambda}) &\approx \frac{1 - \exp\{t\theta E[S_N | S_N \leq 0, \hat{\lambda}]\}}{e^{ht\theta} \exp\{t\theta E[S_N - h | S_N \geq h, \hat{\lambda}]\} - \exp\{t\theta E[S_N | S_N \leq 0, \hat{\lambda}]\}} \\
 &\approx \frac{1 - e^{t\theta\rho_-}}{e^{t\theta(h+\rho_+)} - e^{t\theta\rho_-}} \\
 &= \frac{e^{-t\theta\rho_-} - 1}{e^{t\theta(h+\rho_+-\rho_-)} - 1} \\
 &\approx \frac{t\theta\rho_-}{1 - e^{t\theta(h+\rho_+-\rho_-)}} . \quad (54)
 \end{aligned}$$

Substituting equations (47), (48), and (54) into equations (40), (41), and (42), respectively, followed by substitution into equation (39), results in the Siegmund-based approximation to the ASN of the Page test with nuisance parameter estimation,

$$A_S(\theta) = E_{\hat{\lambda}} \left[\frac{h + \rho_+ - \rho_-}{E[g_{\hat{\lambda}}(x)|\hat{\lambda}]} \right] + \frac{E_{\hat{\lambda}} \left[\frac{\rho_-}{E[g_{\hat{\lambda}}(x)|\hat{\lambda}]} \right]}{E_{\hat{\lambda}} \left[\frac{t\theta\rho_-}{1 - e^{t\theta(h+\rho_+-\rho_-)}} \right]} . \quad (55)$$

The corresponding Wald-based approximation ignores the excess over the boundary; that is, $\rho_+ \rightarrow 0$ and $\rho_- \rightarrow 0$. Applying this to equation (55) and utilizing L'Hospital's rule yields

$$\begin{aligned}
A_W(\theta) &= \lim_{\rho_{\pm} \rightarrow 0} A_S(\theta) \\
&= E_{\hat{\lambda}} \left[\frac{h}{E[g_{\hat{\lambda}}(x) | \hat{\lambda}]} \right] + \lim_{\rho_{\pm} \rightarrow 0} \frac{E_{\hat{\lambda}} \left[\frac{\rho_-}{E[g_{\hat{\lambda}}(x) | \hat{\lambda}]} \right]}{E_{\hat{\lambda}} \left[\frac{t_{\theta} \rho_-}{1 - e^{t_{\theta}(h + \rho_+ - \rho_-)}} \right]} \\
&= E_{\hat{\lambda}} \left[\frac{h}{E[g_{\hat{\lambda}}(x) | \hat{\lambda}]} \right] + \lim_{\rho_- \rightarrow 0} \frac{E_{\hat{\lambda}} \left[\frac{1}{E[g_{\hat{\lambda}}(x) | \hat{\lambda}]} \right]}{E_{\hat{\lambda}} \left[\frac{t_{\theta} \{1 - (1 + t_{\theta} \rho_-) e^{t_{\theta}(h - \rho_-)}\}}{(1 - e^{t_{\theta}(h - \rho_-)})^2} \right]} \\
&= E_{\hat{\lambda}} \left[E[g_{\hat{\lambda}}(x) | \hat{\lambda}]^{-1} \right] \left\{ h + E_{\hat{\lambda}} \left[\frac{t_{\theta}}{1 - e^{t_{\theta} h}} \right]^{-1} \right\}. \tag{56}
\end{aligned}$$

Thus, equations (55) and (56) describe the Siegmund- and Wald-based approximations to the ASN of the Page test with nuisance parameter estimation. The MGF root t_{θ} , Siegmund corrections ρ_+ and ρ_- , and nonlinearity mean $E[g_{\hat{\lambda}}(x)]$ must be determined on a case-by-case basis as functions of the NPE $\hat{\lambda}$ and substituted into these equations. Under special circumstances (e.g., $t_{\theta_1} < 0$ and $t_{\theta_0} > 0$ for all $\hat{\lambda}$) these equations may yield simpler forms accurate for approximation to \bar{T} and \bar{D} .

Relationship to Standard Page Test

If the amount of data used to estimate the nuisance parameter λ tends to infinity and the NPE $\hat{\lambda}$ is consistent, $\hat{\lambda} \rightarrow \lambda$ and the Page test with nuisance parameter estimation becomes the standard Page test. Thus, the expectations in equations (55) and (56) over $\hat{\lambda}$ are removed, which, as expected, results in the standard Siegmund-based approximation

$$A_S(\theta) \rightarrow \frac{1 + t_{\theta}(h + \rho_+ - \rho_-) - e^{t_{\theta}(h + \rho_+ - \rho_-)}}{t_{\theta} E[g_{\lambda}(x)]} \tag{57}$$

and the standard Wald-based approximation

$$A_W(\theta) \rightarrow \frac{1 + t_{\theta} h - e^{t_{\theta} h}}{t_{\theta} E[g_{\lambda}(x)]}. \tag{58}$$

GAUSSIAN SHIFT IN MEAN SIGNAL WITH UNKNOWN VARIANCE

As previously mentioned, a signal type commonly encountered in sonar signal processing may be modeled by a change in the mean of a Gaussian random variable. Additionally, suppose that the variance of the Gaussian random variable is unknown but constant. Under these conditions, the observed data are distributed according to

$$x_i \sim \begin{cases} \mathcal{N}(0, \sigma^2) & i < p \\ \mathcal{N}(\mu, \sigma^2) & i \geq p \end{cases}, \quad (59)$$

where p is the unknown starting time of the signal, the mean μ is known, and the variance σ^2 is unknown and considered to be a nuisance parameter. If the auxiliary data, say $\{Y_j\}_{j=1}^J$, are assumed to be signal-free, the MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{J} \sum_{j=1}^J Y_j^2. \quad (60)$$

In this case, the unbiased estimator for the variance under the signal-present hypothesis,

$$\hat{\sigma}^2 = \frac{1}{J} \sum_{j=1}^{J+1} (Y_j - \bar{Y})^2, \quad (61)$$

may be equivalently used, where

$$\bar{Y} = \frac{1}{J+1} \sum_{j=1}^{J+1} Y_j \quad (62)$$

and it is assumed that the estimate is formed from $J+1$ samples of auxiliary data rather than J to retain statistical equivalence between the estimators.

Appropriately scaled, $\hat{\sigma}^2$ (for either of the above estimators) follows a central chi-squared distribution with J degrees of freedom,

$$W = \frac{J\hat{\sigma}^2}{\sigma^2} \sim \chi_J^2. \quad (63)$$

Also note that $\hat{\sigma}^2$ is unbiased for σ^2 ; that is, $E[\hat{\sigma}^2] = \sigma^2$.

The LLR or locally optimal nonlinearity, with the NPE $\hat{\sigma}^2$ substituted for σ^2 , has the form

$$g(x) = \frac{\mu}{\hat{\sigma}^2} \left(x - \frac{\mu}{2} \right), \quad (64)$$

which, in this case, is the actual LLR scaled by $\frac{\sigma^2}{\hat{\sigma}^2}$. The mean, variance, and MGF unity-root of $g(x)$ conditioned on the NPE $\hat{\sigma}^2$ under the signal-present ($\theta = E[x] = \mu$) and signal-absent ($\theta = E[x] = 0$) hypotheses are found in table 4. Note that the MGF unity-roots are simply ± 1 (the MGF unity-roots for the LLR) divided by the scale term, $\frac{\sigma^2}{\hat{\sigma}^2}$.

Table 4. Mean, Variance, and MGF Unity-Root for Gaussian Shift in Mean Signal Conditioned on the NPE $\hat{\sigma}^2$

	Signal-present ($\theta = \mu$)	Signal-absent ($\theta = 0$)
$E_{\theta} [g(x)]$	$\frac{\mu^2}{2\hat{\sigma}^2}$	$\frac{-\mu^2}{2\hat{\sigma}^2}$
$\text{Var}_{\theta} [g(x)]$	$\frac{\mu^2\sigma^2}{\hat{\sigma}^4}$	$\frac{\mu^2\sigma^2}{\hat{\sigma}^4}$
MGF Root t_{θ}	$-\frac{\hat{\sigma}^2}{\sigma^2}$	$\frac{\hat{\sigma}^2}{\sigma^2}$

Conditioned on $\hat{\sigma}^2$, $g(x)$ is still Gaussian under both the signal-present and signal-absent hypotheses. Thus, as found in table 1, the Siegmund correction terms are

$$\rho_+ = 0.5825 \frac{\mu\sigma}{\hat{\sigma}^2} \quad (65)$$

and

$$\rho_- = -\rho_+ = -0.5825 \frac{\mu\sigma}{\hat{\sigma}^2} . \quad (66)$$

Average Delay Before Detection

In order to determine the Siegmund-based average delay before detection, $\bar{D}_S = A_S(\theta_1 = \mu)$, each of the terms in equation (55) must be determined under the signal-present hypothesis. For notational convenience, let $\gamma = 0.5825$. These terms, which are expectations over the NPE (in this case, $\hat{\sigma}^2$), are as follows:

$$E_{\hat{\sigma}^2} \left[\frac{h + \rho_+ - \rho_-}{E[g(x) | \hat{\sigma}^2]} \right] = E_{\hat{\sigma}^2} \left[\frac{2\hat{\sigma}^2}{\mu^2} \left(h + 2\gamma \frac{\mu\sigma}{\hat{\sigma}^2} \right) \right]$$

$$\begin{aligned}
&= \frac{2}{\mu^2} (h E_{\hat{\sigma}^2} [\hat{\sigma}^2] + 2\gamma\mu\sigma) \\
&= \frac{2}{\mu^2} (h\sigma^2 + 2\gamma\mu\sigma), \tag{67}
\end{aligned}$$

$$\begin{aligned}
E_{\hat{\sigma}^2} \left[\frac{\rho_-}{E[g(x)|\hat{\sigma}^2]} \right] &= E_{\hat{\sigma}^2} \left[\frac{2\hat{\sigma}^2}{\mu^2} \left(\frac{-\gamma\mu\sigma}{\hat{\sigma}^2} \right) \right] \\
&= -2\gamma \frac{\sigma}{\mu}, \tag{68}
\end{aligned}$$

and

$$\begin{aligned}
E_{\hat{\sigma}^2} \left[\frac{t_\mu \rho_-}{1 - e^{t_\mu(h+\rho_+ - \rho_-)}} \right] &= E_{\hat{\sigma}^2} \left[\frac{\left(\frac{-\hat{\sigma}^2}{\sigma^2} \right) \left(\frac{-\gamma\mu\sigma}{\hat{\sigma}^2} \right)}{1 - \exp \left\{ -\frac{\hat{\sigma}^2}{\sigma^2} \left(h + 2\gamma \frac{\mu\sigma}{\hat{\sigma}^2} \right) \right\}} \right] \\
&= \gamma \frac{\mu}{\sigma} E_{\hat{\sigma}^2} \left[\frac{1}{1 - \exp \left\{ -2\gamma \frac{\mu}{\sigma} - h \frac{\hat{\sigma}^2}{\sigma^2} \right\}} \right] \\
&= \gamma \frac{\mu}{\sigma} E_{\hat{\sigma}^2} \left[\sum_{k=0}^{\infty} \left(\exp \left\{ -2\gamma \frac{\mu}{\sigma} - h \frac{\hat{\sigma}^2}{\sigma^2} \right\} \right)^k \right] \\
&= \gamma \frac{\mu}{\sigma} \sum_{k=0}^{\infty} e^{-2k\gamma \frac{\mu}{\sigma}} E_{\hat{\sigma}^2} \left[e^{-kh \frac{\hat{\sigma}^2}{\sigma^2}} \right] \\
&= \gamma \frac{\mu}{\sigma} \sum_{k=0}^{\infty} e^{-2k\gamma \frac{\mu}{\sigma}} E_W \left[e^{-\frac{kh}{J} W} \right] \\
&= \gamma \frac{\mu}{\sigma} \sum_{k=0}^{\infty} \frac{e^{-2k\gamma \frac{\mu}{\sigma}}}{\left(1 + 2k \frac{h}{J} \right)^{\frac{J}{2}}} \\
&\approx \gamma \frac{\mu}{\sigma} (1 + \epsilon_K), \tag{69}
\end{aligned}$$

where

$$\epsilon_K = \sum_{k=1}^K \frac{e^{-2k\gamma \frac{\mu}{\sigma}}}{\left(1 + 2k \frac{h}{J} \right)^{\frac{J}{2}}} \tag{70}$$

is used to approximate the infinite summation in the next to last line of equation (69). To arrive at the result of equation (69), the MGF of the chi-squared random variable W ,

$$E_W \left[e^{tW} \right] = (1 - 2t)^{-\frac{J}{2}} \tag{71}$$

is required.

Substitution of equations (67)–(69) into equation (55) results in the Siegmund-based approximation to \bar{D} for the Gaussian shift in mean signal with unknown variance,

$$\begin{aligned}\bar{D}_S &= \frac{2}{\mu^2} (h\sigma^2 + 2\gamma\mu\sigma) - \frac{2\sigma^2}{\mu^2(1 + \epsilon_K)} \\ &= \frac{h + 2\gamma\frac{\mu}{\sigma} - \frac{1}{1 + \epsilon_K}}{\frac{\mu^2}{2\sigma^2}}.\end{aligned}\quad (72)$$

In order to determine the Wald-based approximation to the average delay before detection, $\bar{D}_W = A_W(\theta_1 = \mu)$, the following terms in equation (56) must be determined under the signal-present hypothesis:

$$\begin{aligned}\mathbb{E}_{\hat{\sigma}^2} \left[\mathbb{E} \left[g(x) \mid \hat{\sigma}^2 \right]^{-1} \right] &= \mathbb{E}_{\hat{\sigma}^2} \left[\frac{2\hat{\sigma}^2}{\mu^2} \right] \\ &= \frac{2\sigma^2}{\mu^2}\end{aligned}\quad (73)$$

and

$$\begin{aligned}\mathbb{E}_{\hat{\sigma}^2} \left[\frac{t_\mu}{1 - e^{t_\mu h}} \right] &= \mathbb{E}_{\hat{\sigma}^2} \left[\frac{-\frac{\hat{\sigma}^2}{\sigma^2}}{1 - \exp \left\{ -h \frac{\hat{\sigma}^2}{\sigma^2} \right\}} \right] \\ &= -\frac{1}{J} \mathbb{E}_W \left[\frac{W}{1 - \exp \left\{ -\frac{h}{J} W \right\}} \right] \\ &= -\frac{1}{J} \mathbb{E}_W \left[W \sum_{k=0}^{\infty} e^{-k \frac{h}{J} W} \right] \\ &= -\frac{1}{J} \sum_{k=0}^{\infty} \mathbb{E}_W \left[W e^{-k \frac{h}{J} W} \right] \\ &= -\frac{1}{J} \sum_{k=0}^{\infty} J \left(1 + 2k \frac{h}{J} \right)^{-\frac{J}{2} - 1} \\ &\approx -(1 + \delta_K),\end{aligned}\quad (74)$$

where

$$\delta_K = \sum_{k=1}^K \frac{1}{\left(1 + 2k \frac{h}{J} \right)^{\frac{J}{2} + 1}} \quad (75)$$

is used to approximate the infinite summation in the next to last line of equation (74). To arrive at the result of equation (74), the first derivative of the MGF of the chi-squared random variable W is required,

$$\mathbb{E}_W \left[W e^{tW} \right] = \frac{d}{dt} \mathbb{E}_W \left[e^{tW} \right]$$

$$\begin{aligned}
&= \frac{d}{dt} (1 - 2t)^{-\frac{J}{2}} \\
&= J(1 - 2t)^{-\frac{J}{2}-1} .
\end{aligned} \tag{76}$$

Substitution of equations (73) and (74) into equation (56) results in the Wald-based approximation to \bar{D} for the Gaussian shift in mean signal with unknown variance,

$$\bar{D}_W = \frac{h - \frac{1}{1+\delta_K}}{\left(\frac{\mu^2}{2\sigma^2}\right)} . \tag{77}$$

The results of equations (72) and (77) are compared to the simulated average delay before detection as a function of SNR in figure 9 for $J = 20$ and $h = 5$. The simulation used 1000 trials that were terminated if a threshold crossing did not occur within 1000 samples (all trials terminated prior to this limit). Throughout this and the following sections, the ϵ_K and δ_K summations in the Siegmund and Wald approximations are terminated at $K = 10$ and the variance estimator of equation (61) is used. Similar to the analysis of the Page test when the nuisance parameters are known, the Siegmund-based approximation of equation (72) provided substantially more accurate results than the Wald-based approximation.

Average Time Between False Alarms

In order to determine the Siegmund-based average time between false alarms, $\bar{T}_S = A_S(\theta_0 = 0)$, each term in equation (55) must be determined under the signal-absent hypothesis:

$$\begin{aligned}
\mathbb{E}_{\hat{\sigma}^2} \left[\frac{h + \rho_+ - \rho_-}{\mathbb{E}[g(x) | \hat{\sigma}^2]} \right] &= \mathbb{E}_{\hat{\sigma}^2} \left[-\frac{2\hat{\sigma}^2}{\mu^2} \left(h + 2\gamma \frac{\mu\sigma}{\hat{\sigma}^2} \right) \right] \\
&= -\frac{2}{\mu^2} \left(h \mathbb{E}_{\hat{\sigma}^2} [\hat{\sigma}^2] + 2\gamma\mu\sigma \right) \\
&= -\frac{2}{\mu^2} \left(h\sigma^2 + 2\gamma\mu\sigma \right) ,
\end{aligned} \tag{78}$$

$$\begin{aligned}
\mathbb{E}_{\hat{\sigma}^2} \left[\frac{\rho_-}{\mathbb{E}[g(x) | \hat{\sigma}^2]} \right] &= \mathbb{E}_{\hat{\sigma}^2} \left[\frac{-2\hat{\sigma}^2}{\mu^2} \left(\frac{-\gamma\mu\sigma}{\hat{\sigma}^2} \right) \right] \\
&= 2\gamma \frac{\sigma}{\mu} ,
\end{aligned} \tag{79}$$

and

$$\begin{aligned}
E_{\hat{\sigma}^2} \left[\frac{t_0 \rho_-}{1 - e^{t_0(h + \rho_+ - \rho_-)}} \right] &= E_{\hat{\sigma}^2} \left[\frac{\left(\frac{\hat{\sigma}^2}{\sigma^2}\right) \left(-\frac{\gamma \mu \sigma}{\hat{\sigma}^2}\right)}{1 - \exp \left\{ \frac{\hat{\sigma}^2}{\sigma^2} \left(h + 2\gamma \frac{\mu \sigma}{\hat{\sigma}^2} \right) \right\}} \right] \\
&= \gamma \frac{\mu}{\sigma} E_W \left[\frac{\exp \left\{ -\frac{h}{J} W - 2\gamma \frac{\mu}{\sigma} \right\}}{1 - \exp \left\{ -\frac{h}{J} W - 2\gamma \frac{\mu}{\sigma} \right\}} \right] \\
&= \gamma \frac{\mu}{\sigma} \sum_{k=1}^{\infty} e^{-2k\gamma \frac{\mu}{\sigma}} E_W \left[e^{-k \frac{h}{J} W} \right] \\
&= \gamma \frac{\mu}{\sigma} \sum_{k=1}^{\infty} \frac{e^{-2k\gamma \frac{\mu}{\sigma}}}{\left(1 + 2k \frac{h}{J}\right)^{\frac{J}{2}}} \\
&\approx \gamma \frac{\mu}{\sigma} \epsilon_K.
\end{aligned} \tag{80}$$

Substitution of equations (78)–(80) into equation (55) results in the Siegmund-based approximation to \bar{T} for the Gaussian shift in mean signal with unknown variance,

$$\begin{aligned}
\bar{T}_S &= -\frac{2}{\mu^2} \left(h\sigma^2 + 2\gamma\mu\sigma \right) + \frac{2\sigma^2}{\mu^2 \epsilon_K} \\
&= \frac{h + 2\gamma \frac{\mu}{\sigma} - \frac{1}{\epsilon_K}}{-\frac{\mu^2}{2\sigma^2}}.
\end{aligned} \tag{81}$$

In order to determine the Wald-based approximation to the average time between false alarms, $\bar{T}_W = A_W(\theta_0 = 0)$, each of the terms in equation (56) must be determined under the signal-absent hypothesis:

$$\begin{aligned}
E_{\hat{\sigma}^2} \left[E \left[g(x) \mid \hat{\sigma}^2 \right]^{-1} \right] &= E_{\hat{\sigma}^2} \left[\frac{-2\hat{\sigma}^2}{\mu^2} \right] \\
&= -\frac{2\sigma^2}{\mu^2}
\end{aligned} \tag{82}$$

and

$$\begin{aligned}
E_{\hat{\sigma}^2} \left[\frac{t_0}{1 - e^{t_0 h}} \right] &= E_{\hat{\sigma}^2} \left[\frac{\frac{\hat{\sigma}^2}{\sigma^2}}{1 - \exp \left\{ h \frac{\hat{\sigma}^2}{\sigma^2} \right\}} \right] \\
&= \frac{1}{J} E_W \left[\frac{W}{1 - e^{-\frac{h}{J} W}} \right] \\
&= -\frac{1}{J} E_W \left[\frac{W e^{-\frac{h}{J} W}}{1 - e^{-\frac{h}{J} W}} \right] \\
&= -\frac{1}{J} E_W \left[W \sum_{k=1}^{\infty} e^{-k \frac{h}{J} W} \right]
\end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{J} \sum_{k=1}^{\infty} E_W [W e^{-k \frac{h}{J} W}] \\
&= -\frac{1}{J} \sum_{k=1}^{\infty} J \left(1 + 2k \frac{h}{J}\right)^{-\frac{J}{2}-1} \\
&\approx -\delta_K.
\end{aligned} \tag{83}$$

Substitution of equations (82) and (83) into equation (56) results in the Wald-based approximation to \bar{T} for the Gaussian shift in mean signal with unknown variance,

$$\bar{T}_W = \frac{h - \frac{1}{\delta_K}}{-\frac{\mu^2}{2\sigma^2}}. \tag{84}$$

The results of equations (81) and (84) are compared to the simulated average time between false alarms as a function of threshold in figure 10 for $J = 20$ and a zero dB SNR. The simulation used 5000 trials that were terminated if a threshold crossing did not occur within 2500 samples (two trials failed to terminate prior to this limit). Again, the Siegmund-based approximation of equation (81) provided substantially more accurate results than the Wald-based approximation.

Page Test Performance

The previous two sections have demonstrated the quality of the Siegmund-based approximations to \bar{D} and \bar{T} for the Page test with nuisance parameter estimation. In this section, these approximations are used to investigate the performance as the amount of auxiliary data, the SNR, and the threshold are varied. Figure 11 contains a Page test receiver operating characteristic (a plot of \bar{D} against \bar{T}) where it is seen that the performance tends to that of the standard Page test when the variance of the data is known ($J = \infty$) as the amount of auxiliary data increases. Figure 12 demonstrates how the threshold required to achieve a desired \bar{T} increases as the amount of auxiliary data diminishes, which is explained by the increased variability of the Page test update $g_{\hat{\lambda}}(x_i)$ due to the increased variability of the NPE $\hat{\lambda}$. The loss incurred by having to estimate the unknown variance is seen in figure 13 as an increase in \bar{D} when \bar{T} is held constant.

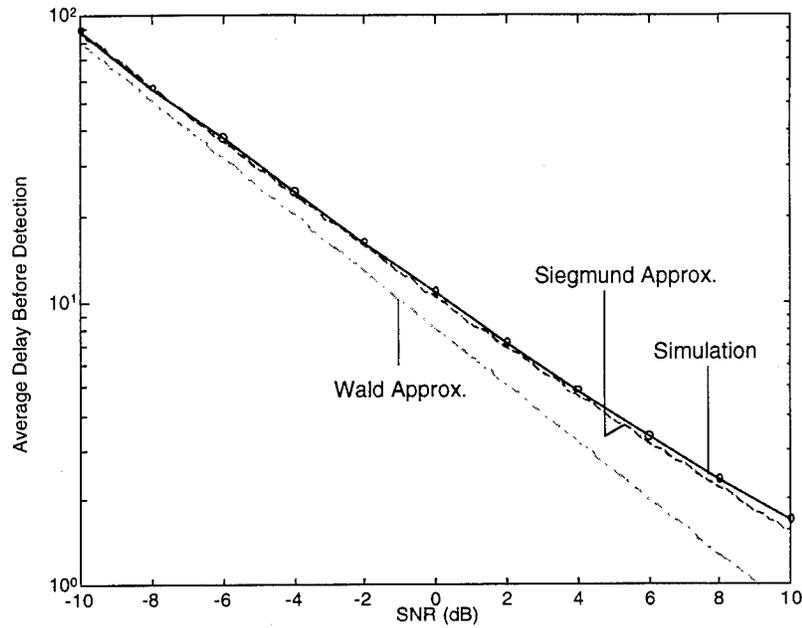


Figure 9. Simulation and Wald and Siegmund Approximations to the Average Delay Before Detection for Gaussian Shift in Mean Signal With Threshold $h = 5$ for Page Test With Nuisance Parameter Estimation

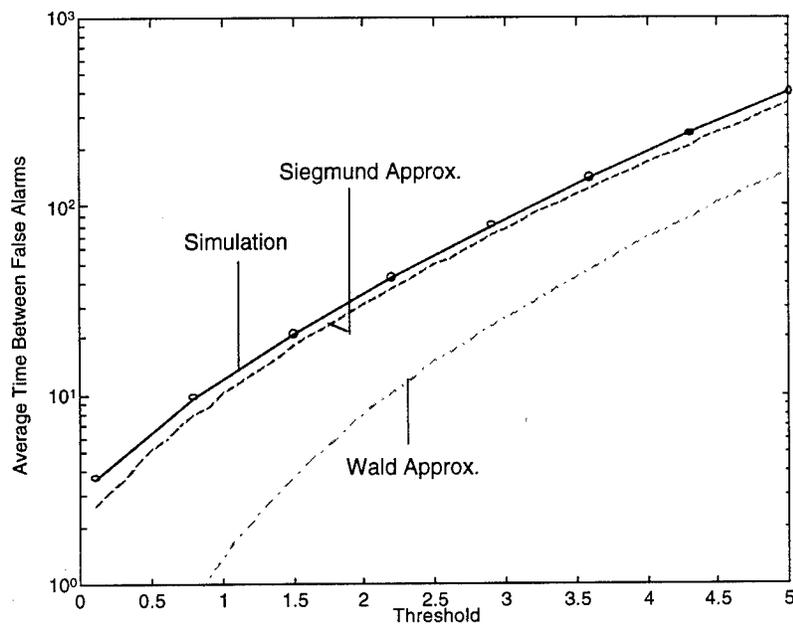


Figure 10. Simulation and Wald and Siegmund Approximations to the Average Time Between False Alarms for Gaussian Shift in Mean Signal for Page Test With Nuisance Parameter Estimation

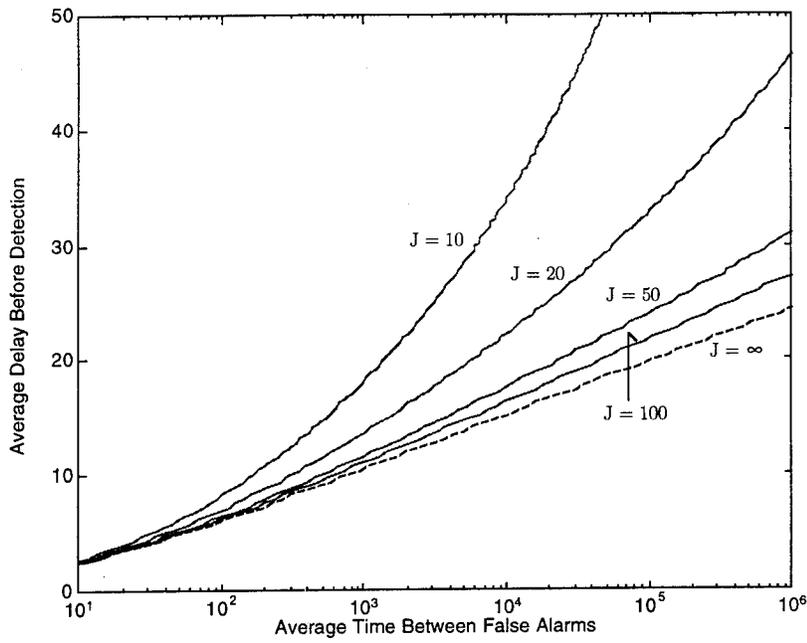


Figure 11. Average Delay Before Detection Vs. Average Time Between False Alarms for 0 dB SNR and Varying Amounts of Auxiliary Data

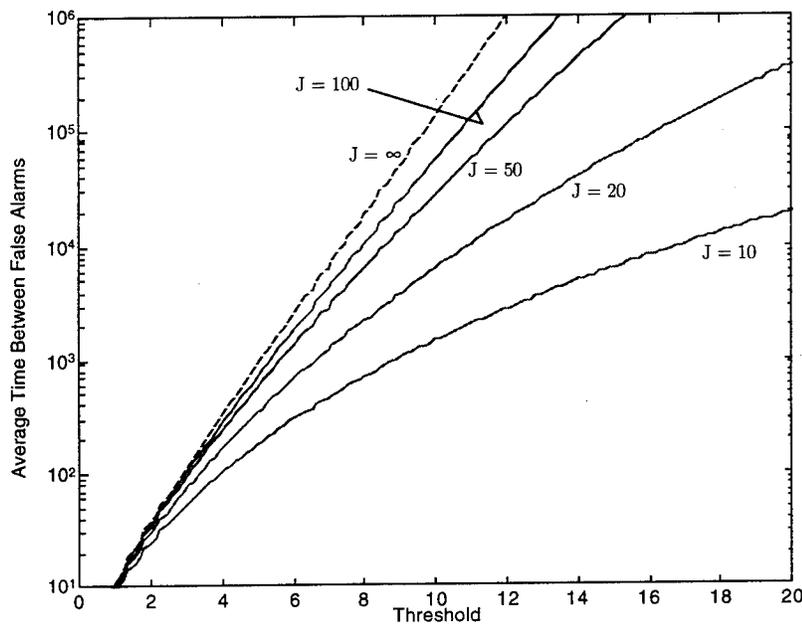


Figure 12. Average Time Between False Alarms Vs. Threshold for 0 dB SNR and Varying Amounts of Auxiliary Data

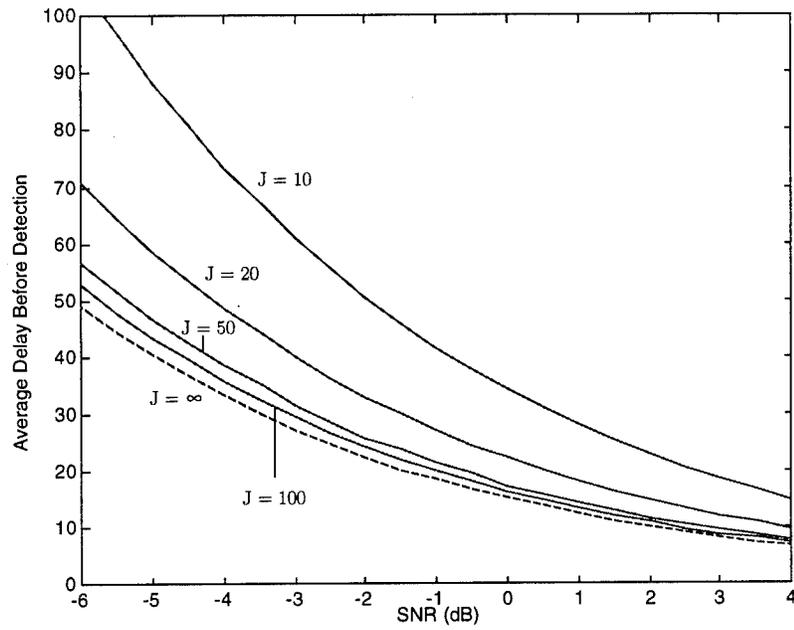


Figure 13. Average Delay Before Detection Vs. SNR for $\bar{T} = 10^4$ Samples and Varying Amounts of Auxiliary Data

Asymptotic Performance

Most applications of the Page test will require a large average time between false alarms. This results in a large threshold, h , which can simplify the computation of the threshold required to achieve a specific \bar{T} and also provide insight into the relationship between \bar{D} and \bar{T} . Assuming that h is large, the Wald approximations to \bar{T} and \bar{D} for the standard Page test, equations (9) and (10), result in the asymptotic approximations

$$\bar{T} \approx \frac{e^{ht_{\theta_0}}}{-t_{\theta_0} \mathbb{E}_{\theta_0} [g(x)]} \quad (85)$$

and

$$\bar{D} \approx \frac{h}{\mathbb{E}_{\theta_1} [g(x)]}, \quad (86)$$

where the conditions $t_{\theta_0} > 0$ and $t_{\theta_1} < 0$ have been exploited. The exponential relationship between h and \bar{T} and the linear relationship between h and \bar{D} are evident from equations (85) and (86).

Broder [16] carried this analysis a step further to demonstrate the asymptotic (large h) exponential relationship between \bar{T} and \bar{D} by introducing an asymptotic performance measure and relating it to the MGF root under the signal-absent hypothesis and the mean of the Page test statistic update under the signal-present hypothesis,

$$\begin{aligned} \eta &= \lim_{h \rightarrow \infty} \frac{\log \bar{T}}{\bar{D}} \\ &= t_{\theta_0} \mathbb{E}_{\theta_1} [g(x)]. \end{aligned} \quad (87)$$

For the Gaussian shift in mean signal where the LLR is used as a nonlinearity, Broder's asymptotic performance measure is

$$\eta = \frac{\mu^2}{2\sigma^2}. \quad (88)$$

Similarly, such asymptotic relationships can be derived for the Page test with nuisance parameter estimation for the Gaussian shift in mean signal. First, observe that δ_K of equation (75) may be approximated by the first term in the infinite summation when h is large enough,

$$\delta_K = \frac{1}{\left(1 + 2\frac{h}{J}\right)^{\frac{J}{2}+1}} + \frac{1}{\left(1 + 4\frac{h}{J}\right)^{\frac{J}{2}+1}} + \frac{1}{\left(1 + 6\frac{h}{J}\right)^{\frac{J}{2}+1}} + \dots$$

$$\approx \frac{1}{\left(1 + 2\frac{h}{J}\right)^{\frac{J}{2}+1}}. \quad (89)$$

Use of this approximation in equations (77) and (84) results in the large threshold approximations

$$\begin{aligned} \bar{D}_W &\approx \frac{h-1}{\eta} \\ &\approx \frac{h}{\eta}, \end{aligned} \quad (90)$$

and

$$\begin{aligned} \bar{T}_W &\approx \frac{h - \left(1 + \frac{2h}{J}\right)^{\frac{J}{2}+1}}{-\eta} \\ &\approx \frac{\left(1 + \frac{2h}{J}\right)^{\frac{J}{2}}}{\eta}, \end{aligned} \quad (91)$$

where η is as in equation (88) and it has been assumed that J is also somewhat large. The linear relationship between h and \bar{D}_W is seen from equation (90). However, the exponential relationship that exists in the standard Page test for the Gaussian shift in mean signal becomes the power law relationship described by equation (91) when the nuisance parameters are estimated. As expected, the power law relationship turns into the exponential one of equation (85) when the amount of auxiliary data increases to infinity ($J \rightarrow \infty$),

$$\bar{T}_W \rightarrow \frac{e^h}{\eta}. \quad (92)$$

This may be seen by noting that $-t_{\theta_0} E_{\theta_0} [g(x)] = \eta$ and $t_{\theta_0} = 1$ for the Gaussian shift in mean signal case using the LLR and by applying the identity

$$\lim_{x \rightarrow \infty} \left(1 + \frac{a}{x}\right)^{bx} = e^{ab}. \quad (93)$$

CONCLUSIONS

This report has presented a method for dealing with unknown noise or nuisance parameters in the Page test for the detection of the onset of a signal. The algorithm is a generalization of the standard Page test where the nuisance parameters are estimated using data previous to a most recent reset—data declared to be signal-free by the Page test. This also makes analysis feasible due to the independence of these data and the data after the most recent reset. The ASNs of the Page test with nuisance parameter estimation were derived using both Wald- and Siegmund-based approximations. It was shown that as the estimation of the nuisance parameter becomes perfect (i.e., infinite auxiliary data), the ASN approximations of the Page test with nuisance parameter estimation simplify to the ASN approximations of the standard Page test.

A Gaussian shift in mean signal with an unknown variance was considered as an example. Closed forms for the Wald- and Siegmund-based approximations to the average time before detection and between false alarms for the Page test with nuisance parameter estimation were derived. The validity of the approximations was verified by comparison to simulation results where it was observed that, as in the standard Page test, the Siegmund-based approximations provide higher accuracy. The Siegmund-based approximations were used to investigate the loss in performance incurred by having to estimate the unknown variance; specifically, the false alarm performance (\bar{T}) as a function of threshold and the detection performance (\bar{D}) as a function of SNR.

The asymptotic, in the sense of a large threshold, relationship between the threshold and the average time before detection for the Page test with nuisance parameter estimation was found to be linear, as with the standard Page test. However, the asymptotic relationship between the threshold and the average time between false alarms was found to follow a power law that approximates the exponential relationship of the standard Page test when the amount of auxiliary data used to estimate the nuisance parameters becomes large.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
Coast Guard Academy (J. Wolcin)	1
Office of Naval Research (Code 451: T. Goldsberry, N. Harned, M. Shipley)	3
Program Executive Office-USW-ASTO (J. Polcari)	1
Space and Naval Warfare Systems Command (R. Holland)	1
Naval Undersea Warfare Center Detachment, West Palm Beach (R. Kennedy)	1
Pennsylvania State University (R. Young)	1
Princeton University (S. Schwartz)	1
University of Connecticut (Y. Bar Shalom, D. Dey, N. Mukhopadhyay, P. Willett)	4
University of Naples (F. Palmieri)	1
University of Rhode Island (D. Tufts)	1
Defense Technical Information Center	12